Visual Analytics for Extracting Trends from Spatio-temporal Data*

Michiel Dhont[1,2][0000−0002−5679−0991], Elena Tsiporkova[1], Tom Tourwé[1], and Nicolás González-Deleito[1]

[1] Sirris, Bd. A. Reyerslaan 80, 1030 Brussels, Belgium
michiel.dhont@sirris.be, elena.tsiporkova@sirris.be, tom.tourwe@sirris.be, nicolas.gonzalez@sirris.be
[2] Department of Electronics and Information Processing (ETRO), VUB, Belgium

Abstract. Visual analytics combines advanced visualisation methods with intelligent analysis techniques in order to explore large data sets whose complexity, underlying structure and inherent dynamics are beyond what traditional visualisation techniques can handle. The ultimate goal is to expose relevant patterns and relationships from the data, since not everything can be exposed easily through intelligent analysis techniques. On the contrary, the human eye can outperform algorithms in grasping and interpreting subtle patterns, provided it is supported by intelligent visualisations.
In this paper, we propose three novel visual analytics techniques for analysing spatio-temporal data. First, we present a fingerprinting technique for discovering and rapidly interpreting temporal and recurring patterns by use of circular heat maps. Next, we present a technique supporting comparisons in time or space by use of circular heat map subtraction. Finally, we propose a technique enabling to characterise and get insights of the temporal behaviour of the phenomenon under study by use of label maps.
The potential of the proposed approach to reveal interesting patterns is demonstrated in a case study using traffic data, originating from multiple inductive loops in the Brussels-Capital Region, Belgium.

Keywords: Visual Analytics · Temporal Statistical Analysis · Spatio-temporal Clustering · Traffic · Covid-19.

1 Introduction

Although the recent advances in AI and in particular deep learning techniques made the exploitation of large volumes of data widely accessible, their black box nature does not offer an intuitive way of getting a deeper understanding of the underlying mechanisms of the phenomenon under study. This is usually pursued via statistical analysis and visualisation techniques, such as correlation analysis,

*This research was supported by the Brussels-Capital Region - Innoviris. This research received funding from the Flemish Government (AI Research Program).

histograms, box- and time-plots, etc. However, such traditional techniques are rather limited when dealing with complex data sets consisting of multivariate spatio-temporal data covering large periods of time.

Visual analytics combines advanced visualisation methods with intelligent analysis techniques in order to explore large data sets whose complexity, underlying structure and inherent dynamics are beyond what traditional visualisation techniques can handle. The ultimate goal is to expose relevant patterns and relationships from the data since not everything can be revealed easily through intelligent analysis techniques. On the contrary, the human eye can outperform algorithms in grasping and interpreting subtle patterns, provided it is supported by intelligent visualisations. Complex data sets need therefore to be manipulated in an intelligent way in order to reveal and highlight the underlying patterns and relationships. This is exactly what visual analytics is about, i.e. advanced data analysis techniques combined with (interactive) visualisation algorithms in order to support the analytical reasoning for decision making.

This paper is concerned with novel visual analytics techniques for spatio-temporal data. First, we present a fingerprinting technique for discovering temporal and recurring patterns by use of circular heat maps. Next, we present a technique supporting comparisons in time or space by use of circular heat map subtraction. Circular heat maps facilitate rapid identification and interpretation of temporal patterns. Finally, we propose a technique enabling to characterise and get insights of the temporal behaviour of the phenomenon under study by use of label maps.

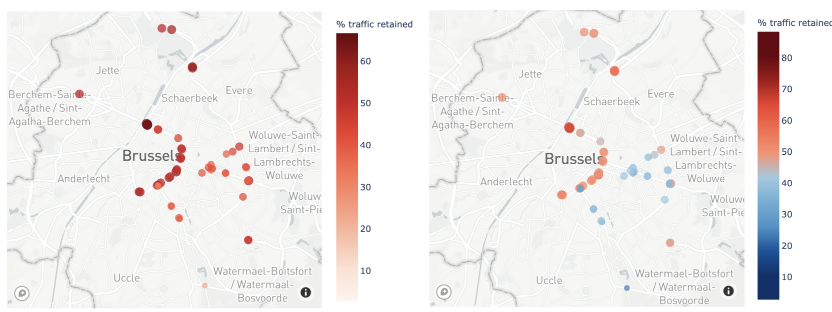
The techniques are application-agnostic and can be used for exploring and extracting relevant insights from spatio-temporal data in domains as traffic, solar and wind energy, electricity consumption, etc. The potential of the proposed approach to reveal interesting patterns is illustrated on recent publicly available traffic count data from the Brussels-Capital Region in Belgium, including the lockdown period imposed by the Covid-19 measures. The latter offers more opportunities for discovering some intriguing trends in the otherwise pretty monotone and typically dense Brussels's traffic.

The rest of the paper is organised as follows. Section 2 presents a literature study of related work, together with a motivation of the rationale of this work. Both the used and novel proposed methods are explained in Section 3. Section 4 illustrates the proposed methods on a case study of Brussels traffic data. Finally, Section 5 concludes the paper with some possible extensions for further research.

2 Rationale and Related Work

2.1 Rationale

In many application domains, data sets explicitly include a location component, often next to a temporal component. Examples of such spatio-temporal data sets can be found in the renewable energy, mobility or environmental domains. Extracting trends and insights from such data sets by using statistical methods only is not sufficient, as these typically do not exploit the spatial information, nor correlate it with the temporal information. For this reason, visual analytics can serve as the key means to support users in exploring aspects of interest within such data. The use of well-thought visualisations (think of aspects as colours, shapes, positioning, etc.), of suitably processed data (normalisation, aggregation, data imputation, clustering, etc.), explicitly linking the spatial and temporal information, enables the user to derive insights beyond what standard statistical methods can achieve. It optimally supports the competences of the human eye to detect and interpret visual structures, and is hence instrumental in more advanced exploitation of such data, as it facilitates the understanding of the underlying mechanisms of the phenomenon under study.



(a) Naive approach (b) Smart colour usage

Fig. 1: Two ways to visualise retained Brussels traffic by Covid-19 restrictions

We illustrate this with a real-life example. In January 2020, we started gathering publicly available traffic data captured at 55 locations in the Brussels-Capital Region, Belgium, without having the slightest suspicion that the escalation of the Covid-19 pandemic would deliver interesting data to analyse. The lockdown measures introduced on March 13$^{\text{th}}$ imposed serious restrictions on traffic, permitting only limited commuting. As expected, the overall traffic volume reduced dramatically during this period. Figure 1a and Figure 1b illustrate the percentage of traffic volume retained during the first 4 weeks of the lockdown in comparison to the period before, for all the observed locations. In these figures, the higher the colour intensity of the small circles denoting the different locations, the higher the percentage of regular traffic that was retained for this location. In the naive approach of Figure 1a, only tints of red are used, which does not reveal much. However, using a colour range with two colours as in Figure 1b allows to zoom in deeper in the data. In that figure, the border between blue and red is fixed on the mean retained traffic over all the locations on the map. In this way, the blue circles denote locations where a bigger reduction of traffic was observed during the lockdown restrictions. One can now clearly see that the ring road around Brussels's city centre retains proportionally more traffic volume than the residential areas around it. This means that the functional traffic in the Brussels-Capital Region has been less impacted by the Covid-19 restrictions than the recreational traffic. This example illustrates that even basic data analysis can provide more value to existing visualisation methods.

2.2 Related Work

Visual Analytics The availability of large amounts of data and the ability to analyse and understand it is becoming ever more relevant and important. By automatically exposing the underlying information, via advanced data analysis, one is able to take much more informed decisions. This is useful in some domains, such as marketing, but vital in the field of medical research and political decisions. For instance, the National Visualisation and Analytics Centre (NVAC) from the United States has the mission to use next-generation technologies to reduce the risk of terrorism. In 2006, NVAC assembled a panel of about 40 leading experts from government agencies, industry and academia to outline an R&D agenda [14] with the explicit goal to advance the state of the science to enable analysts to detect the expected and discover the unexpected from massive and dynamic information streams and databases consisting of data of multiple types and from multiple sources, even though the data is often conflicting and incomplete. The agenda also defined visual analytics as a multidisciplinary field that includes the following focus areas: (i) analytical reasoning techniques, (ii) visual representations and interaction techniques, (iii) data representations and transformations, (iv) techniques to support production, presentation, and dissemination of analytical results.

Statistical Visualisations Data visualisation is typically concerned with depicting some statistics. There is a wide range of classical approaches available, going from very basic ones (e.g. scatter plots) to complex multi-plot visualisations. Depending on the application context and the data analytics workflow considered, suitable visualisations can be selected [17]. However, researchers should not refrain from experimenting beyond the traditional approaches by combining multiple known visualisations into one comprehensive plot. For instance, Allen et al. [1] proposed to augment a violin plot with scatter plots and similar statistics as in a box plot. Furthermore, they proposed possible extensions on their so called rainbow plots by changing orientations and dividing the data in separate groups. These plots succeed in bringing many pieces of information together in an orderly manner in one visual. Another example is shown in the research of Zhao et al. [18], where the advantage of integrating classical line charts into circular heat maps is illustrated.

Constructing appropriate visualisations can be hard and time consuming. For this reason, research has been conducted on constructing powerful recommendation tools which guide users into obtaining relevant visualisations for their specific analytical task [16,8]. Although promising results have been obtained, the main difficulty resides in transferring the specific intent of the research to the recommendation tool.

Visualisation of Spatio-temporal Data Sensors which capture data at a fixed frequency are omnipresent. To exploit the (multivariate) time-sensitive data they generate, the time aspect is often treated similar as the sensor measures themselves. However, time requires a special treatment since it is not simply a measure. For this purpose, a range of special-purpose time-sensitive visualisation techniques have been developed and proven to be effective. To decide which technique is useful in a certain situation, one needs to consider aspects as whether the data is dynamic, consists out of events, is multivariate, etc. [11]

Real-world time series data is often enriched with position information, but visualising such data is hard, e.g. visualising dynamically changing data across different geographical locations. Rodrigues et al. [10] proposed a basic two-tier interface to tackle such challenges, which they validated on data concerning energy production of power plants. In the first tier, users see a geographical map indicating information with only few details. To access the second tier, users can select a location, resulting in charts of the energy production.

Spatio-temporal data can also be found in the mobility domain. Tang et al. [13] proposed a method to extract, by an interactive visual analysis system, characteristics on specific areas based on GPS data originating from taxis. They used maps to visualise the main traffic flows and heat maps to observe traffic distribution over time. The resulting visualised characteristics can e.g. assist the business development process in choosing locations for new stores.

In [18], Zhao et al. illustrated the convenience of multiple variants on circular heat maps in spatio-temporal data sets. Sun et al. [12] developed a method to embed spatio-temporal information in a map, by the use of on bidirectional line charts on road sections. Andrienko and Andrienko [2] investigated aggregation strategies in case of spatio-temporal data for both traffic-oriented and movement-oriented visualisations. As visualisations, they proposed multiple variants of directed graphs and heat maps.

3 Methods

This section describes the methods used and proposed in this paper. First of all, Section 3.1 describes how cluster analysis is typically performed, as clustering is used in Section 3.4 and later in Section 4. The remaining subsections describe the visual analytics methods proposed in this paper: temporal fingerprinting through circular heat maps (Section 3.2), spatio-temporal comparison through circular heat map subtraction (Section 3.3), and temporal behaviour characterisation through label maps (Section 3.4).

3.1 Cluster Analysis

Clustering approaches are often used in data science to gain valuable insights by observing which data objects are grouped together. To divide data objects into disjoint clusters, the most commonly used partitioning algorithms require that the number of clusters (k) is determined, either beforehand or when determining the best cut in the dendrogram in case of hierarchical clustering [9]. This represents a challenge, since there is often a lack of prior knowledge to decide this number. Determining a correct, or suitable, k is a hard problem in a real-world data set. To address this issue, researchers usually generate clustering results for multiple values of k, and subsequently assess the quality of the obtained clustering solutions. In situations where no prior knowledge is available, assessing the quality of these solutions can be done using several measures, related to e.g. the compactness and separation properties of the solution (*Davis-Bouldin Index* [4]), the connectedness (*Connectivity* [5]), or the ratio of the within-cluster variance with the overall-between cluster variance (*Calinski-Harabasz Index* [3]). In practice, a majority voting approach is often used, combining the results of multiple such validation measures to identify the most optimal number of clusters.

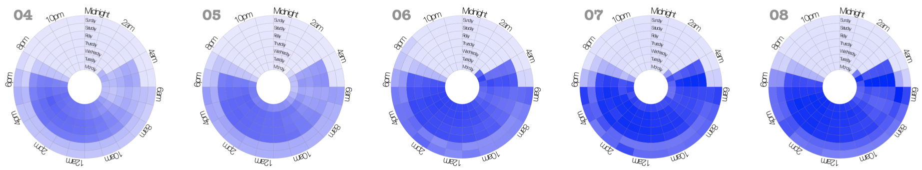
3.2 Temporal Fingerprinting through Circular Heat Maps

The analysis of temporal data can often benefit from comparison between recurring time periods as days, weeks, months, etc. which allows to identify trends and seasonality. This is particularly relevant for applications where the monitored phenomenon can be naturally divided in such periods e.g. periodic electricity consumption of households, traffic intensity, yield of photovoltaic (PV) plants, production efficiency of different shifts in a factory, etc.

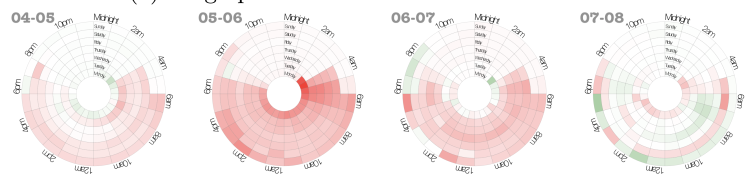
In order to visualise such trends and patterns explicitly, we have developed a general methodology allowing to convert time series data into a series of circular heat maps covering recurring time periods. Each heat map can be interpreted as a characteristic fingerprint facilitating rapid perception of the behaviour of the phenomenon under study for the time period covered by the heat map. The choice of a circular heat map, instead of a simple line chart or a classical rectangular heat map for example, is motivated by its ability of depicting several dimensions or views therein (i.e. days of the week, hours of the day, vehicle counts) in a visually very compact fashion. This compact representation enables a viewer to quickly find patterns in the data without requiring to focus on different potentially far apart points in the figure. Its circular nature also enables to easily highlight patterns occurring at the limits of the circular dimension (e.g. around midnight). Furthermore, through the use of the *small multiples* visualisation technique [15] —a set of similar thumbnail-sized figures which represent the same phenomenon along a different partitioning of the data—, they facilitate comparisons and the highlight of differences. More precisely, by constructing a small multiple from a collection of fingerprints one is able to do a comparison between fingerprints in different time periods or other multiple phenomena (e.g. occurring in different spatial locations).

In Figure 2a, such a circular heat map small multiple is generated for depicting the electricity consumption of a university building in Arizona for 5 consecutive months. In this example, each circular heat map depicts the days of the

[3] Data source: energy consumption at Arizona State University (ASU), https://www.kaggle.com/pdnartreb/asu-buildings-energy-consumption/activity



(a) Fingerprints of consecutive months in 2016



(b) Diff. between consecutive months (red: increase, green: decrease)

Fig. 2: Energy consumption fingerprints for a university building[3]

week as concentric circles starting with Monday in the inner circle, followed by Tuesday in the next circle and so on until placing Sunday in the outermost circle. The circles are divided in 24 sectors of 1 hour, ordered clockwise and starting with midnight at 12 am. The colour of a sector indicates the observed consumption, the darker the colour the higher the observed consumption. In the leftmost fingerprint (April), one can observe that the highest energy consumption occurs between 4 am and 7 pm during weekdays. Thanks to the small multiple, we can observe that the overall consumption pattern is consistent across the months, but the actual consumption increases steadily.

3.3 Spatio-temporal Comparison through Circular Heat Map Subtraction

Detecting Differences in Time The use of circular heat maps and small multiples as illustrated in Figure 2a is well-suited for monitoring evolution in time. One can go even further and subtract the values from two heat maps in order to reveal and highlight better their differences. Subsequently, a heat map with one colour (e.g white) representing identical values and two diverging colours representing the positive and negative differences can offer a very insightful view.

Depending on the application context, two different subtracting approaches can be considered. For both of them, let us consider a sequence of heat maps covering the same time duration (e.g. a week or a month).

– Compare each heat map in the sequence with the heat map from the **previous period** (by subtracting the latter from the former). For this, it is essential that the heat maps are ordered chronologically in time. In this way, by explicitly highlighting the differences between consecutive time periods, one can more clearly observe local changes. For instance, the fingerprints of Figure 2a are subtracted from each other, resulting in Figure 2b. Note that the clear increase of daily consumption between May, June and July and the slight decrease between July and August can be spotted immediately in Figure 2b, which is much less obvious in Figure 2a.

- Compare each heat map in the sequence with a **reference (baseline) heat map** (again, by subtracting the latter from the former) to detect global differences in performance. In this way, by examining the resulting sequence of multiple subtracted heat maps from different time periods (e.g. each week), one can quickly identify which period deviates the most from the expected pattern. Figure 3a shows as baseline the average weekly energy consumption pattern of the university building over the full time period. Figure 3b provides the quarterly differences w.r.t. this baseline as a sequence of subtracted heat maps. The observed increase in energy during Q2 and Q3 could be due to the use of air conditioning during warmer months.

(a) Baseline (b) Quarterly differences (red: increase, green: decrease)

Fig. 3: Quarterly differences of energy consumption, compared to the overall averaged baseline, for a university building

Detecting Differences in Space The previous examples strongly emphasise the time aspect. However, our fingerprinting approach can also be applied for analysing data along the spatial component. For example, it is possible to capture temporal patterns (e.g. weekly electricity consumption) in a representative heat map and subsequently, link multiple locations together into the spatial dimension by constructing a sequence of heat maps covering the same time period for all the considered locations. Note that this is another example of the use of small multiples, and many different variables could be used to construct it (e.g. PV power production per weather condition).

Figure 4 provides the fingerprints of the electricity consumption in February 2016 for 4 different buildings of Arizona State University[3]. One can clearly observe 4 quite different consumption patterns: a more intense consumption occurring during the day on weekdays, esp. during afternoons (heat map labelled with an A); a higher consumption during evenings, both during weekdays and weekends (heat map B); a slightly higher consumption at night and on Mondays (heat map C); and a high consumption spread over all days and hours, but more intense during evenings and at night (heat map D).

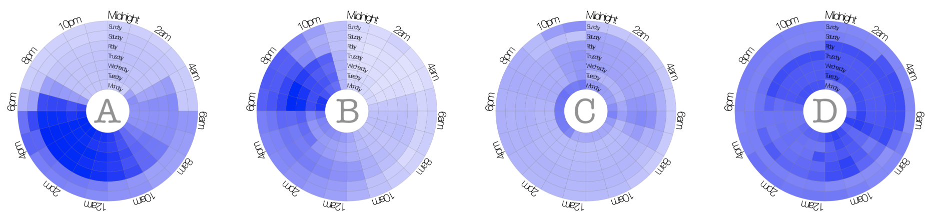


Fig. 4: Electricity consumption of 4 university buildings in February 2016

3.4 Temporal Behaviour Characterisation through Label Maps

The temporal behaviour of a complex phenomenon often consists of a limited set of distinct characteristic profiles, e.g. a wind turbine goes through different operating modes or traffic undergoes peak and off-peak periods. Such characteristic profiles can be extracted from the data using clustering techniques, as shown in the work of Iverson [6] on inductive system health monitoring. It models the relationship between the different variables by considering whether or not their values are sufficiently similar, independent of the temporal component. This results in a limited number of groups that characterise regular but different behaviour. As a result, each timestamp is assigned a particular label that corresponds to this characteristic behaviour profile.

In a second step, such regular behaviour profiles can be used for different purposes e.g. rapid annotation, detecting deviations, understanding state transitions in time, etc. To this extent, we propose to visualise the resulting behaviour characterisation using a label map, a matrix-like visualisation where each column represents a fine-grained view on the time dimension (e.g. a timestamp such as hour of the day), each row a coarser-grained view on time encompassing all the columns of the row (e.g. a day), and each cell is painted with the colour of the profile to which each timestamp belongs. The choice of such a visualisation, instead of a line chart or a circular heat map for example, is motivated by its ability of clearly depicting behaviour expected to be seen in a recurrent fashion (e.g. every day) on the different rows of the matrix. Furthermore, by just adding new rows, the visualisation can be directly used for real-time monitoring.

Figure 5 illustrates a label map in the previously introduced context of the electricity consumption data of a university building. In this visualisation, columns depict timestamps during one day and rows depict individual days. Through clustering, five behaviour profiles have been identified. By looking at when the clusters occur, we can observe a (to be expected) clear repetitive weekly

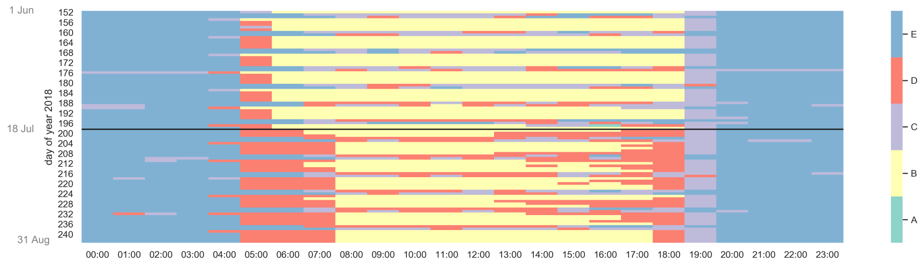


Fig. 5: Evolution of electricity consumption modes in a university building

pattern, consisting of low consumption during night and weekend (blue/E), high consumption during working hours (yellow/B), and three other profiles most probably related to maintenance operations. The visualisation reveals a change starting on July 18^{th}, where the extent of cluster B (yellow) is reduced; this might be due to the start of the summer holidays.

4 Case Study on Brussels Traffic

In this section we illustrate how the proposed methods from Section 3 can be used to derive relevant insights from spatio-temporal data.

4.1 Data

Our analysis is based on open traffic data from the Brussels-Capital Region, Belgium. We started gathering the data in the beginning of 2020. The lockdown measures introduced on March 13^{th} imposed serious restrictions on the traffic in Brussels, permitting only limited commuting. In this way, the observed traffic after the introduction of the measures can be considered as an opportunity to derive some characteristic blueprints of the traffic in Brussels. Since the second half of May, the Covid-19 restrictions were gradually being relaxed, traffic was slowly returning to 'normal' and allowed us to observe the emergence of traffic volumes associated to different activities.

The data contains vehicle counts, average speed measurements and occupancy (percentage of road covered) of 55 busy locations in Brussels. The locations can be observed in Figure 1. Each location represents one direction of a road and combines the information of all available lanes in that direction. The obtained data has a one minute granularity and originates from both ANPR cameras and inductive loops. We collected the data in real-time from a publicly available API of Brussels Mobility[4] over a time span from mid January 2020 until the first week of June 2020.

[4]https://data-mobility.brussels/traffic/api/counts/

4.2 Unravelling Volume Patterns of Brussels Traffic

Weekly Traffic Intensity Patterns City traffic is strongly dependent on the day of the week. Therefore, the fingerprinting approach proposed in Section 3.2 can be applied by segmenting the data per week. In this way, a weekly traffic intensity fingerprint can be extracted for each monitored location in the form of a circular heat map. Like before, we depict in the circular heat maps the days of the week as concentric circles starting with Monday in the inner circle, followed by Tuesday in the next circle and so on until placing Sunday in the outermost circle. The circles are divided in 24 sectors of 1 hour (i.e. vehicle counts are aggregated per hour), ordered clockwise starting with midnight at 12 am.

Such a representation allows to easily compare weekly patterns of different locations for different time periods. For instance, Figure 6 depicts the typical weekly patterns derived for 4 different locations in Brussels for 2 different periods before and during the Covid-19 lockdown (i.e. the corresponding weeks are aggregated into one weekly pattern per location). The same colour scale is used for all the four locations, which facilitates objective comparison across them. It is clear that Troontunnel has more traffic than Belliardtunnel, which has on its turn more traffic than Keizer Karellaan and Vleurgattunnel. This pattern is apparent before as well as during the lockdown. It is also interesting to zoom into the specificity of the weekly traffic behaviour. For instance, despite the well manifested difference in the overall traffic intensity between Vleurgattunnel and Belliardtunnel, a very clear morning peak (dark sector between 7 and 9 am) can be detected for both locations during the working days in the pre-Covid-19 period, while the evening peak (dark sector between 4 and 6 pm) is well established only for Belliardtunnel. During the lockdown, this morning peak during working days is still observable (except for Vleurgattunnel), indicating that some work commuting was still happening through those locations. The third row of the figure depicts the difference between the pre-Covid-19 situation of the top row and the full lockdown situation of the second row. This allows us to observe that night traffic during the weekend days has disappeared completely (esp. visible for the Troontunnel and the Belliardtunnel), that the weekend days had a larger reduction in traffic compared to weekdays for the Troontunnel, and that the reduction was also stronger for rush hours in both Troontunnel and Belliardtunnel.

It is also interesting to study the traffic intensity evolution. In Figure 7, characteristic weekly fingerprints are depicted based on aggregating vehicle counts over all available locations from the 10^{th} until the 15^{th} week of 2020. The week number of each fingerprint is shown in its left upper corner. By examining these fingerprints one can detect easily the introduction of the first Covid-19 restrictions on the Saturday of week 11, and the introduction of the full lockdown on Wednesday at 12 am in week 12. Moreover, by comparing week 10 (pre-Covid-19) with weeks 13, 14 and 15 (full lockdown), it is interesting to observe that the biggest reduction in traffic is obtained during the weekend. The intensity of colours during weeks 13, 14 and 15 is also very similar, indicating that people consistently obeyed the imposed restrictions.

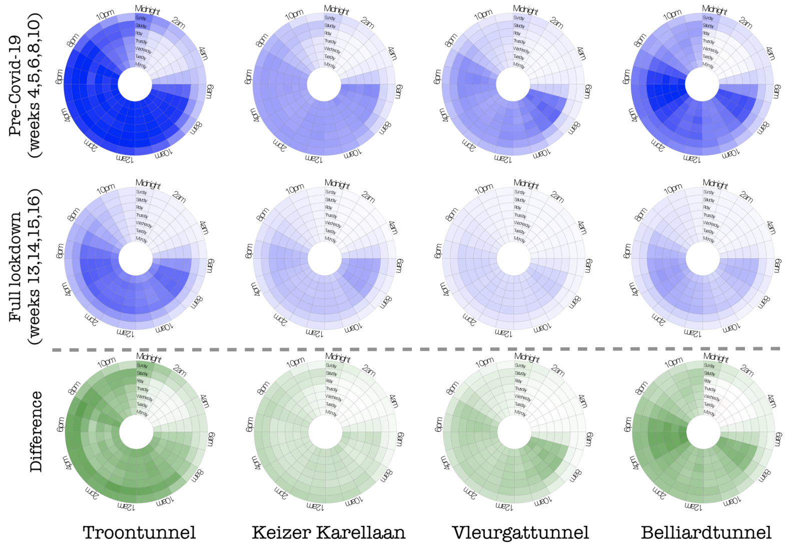


Fig. 6: Vehicle counts before and during the Covid-19 restrictions at 4 locations

Traffic Volume Disaggregation The collected data spans over a time period covering 3 distinct traffic situations: 1) normal traffic referring to regular work-school weeks; 2) carnival holidays referring to the school vacation in week 9, which excludes school-related traffic and some work-related traffic due to parents taking vacation—at the same time, those families have more time for recreational trips (e.g. city trips, sport, shopping, etc.); and 3) lockdown weeks referring to the period of activity restrictions due to the Covid-19 measures, including only work related traffic which cannot be performed via teleworking and other minimal essential traffic (e.g. shopping for food).

The Covid-19 restrictions were gradually relaxed since the second half of May. Traffic was slowly returning to 'normal' and allowed us to observe the emergence of traffic volumes associated to different activities. Following the approach described above, we have generated characteristic weekly fingerprints for different time periods depicted in the upper row of Figure 8. Each of these fingerprints depicts the hourly traffic intensity throughout the (averaged) week. Our baseline (100% traffic volume) is constructed by averaging over 5 'regular' work-school

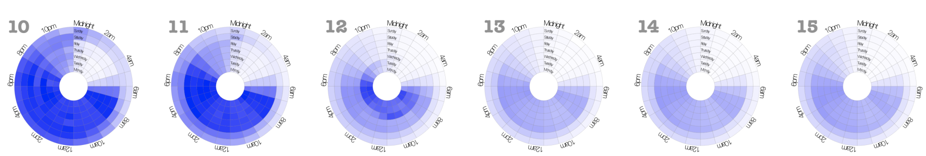


Fig. 7: Weekly evolution of vehicle counts in Brussels (week 10 until week 15)

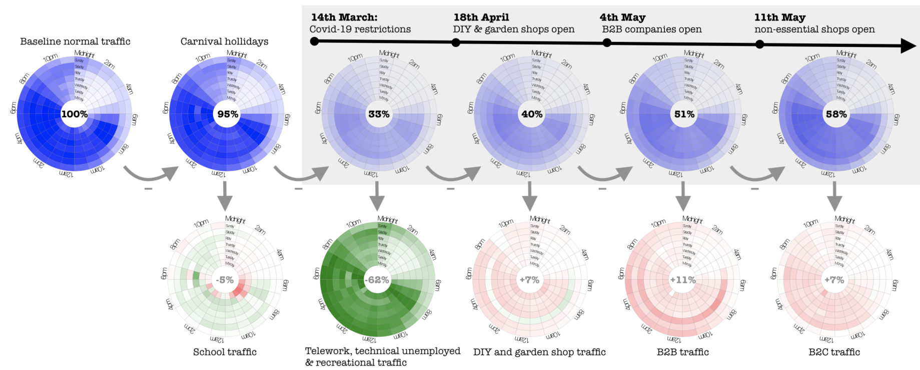


Fig. 8: Upper row: grouped fingerprints of vehicle counts in Brussels. Bottom row: disaggregated fingerprints from above (green: reduction, red: increase)

weeks, excluding school holidays, in January and February 2020. The second fingerprint (Carnival holidays) refers to the school vacation in week 9, while the remaining fingerprints average traffic intensities over the different phases of the lockdown period, i.e. complete lockdown between March 14$^{\text{th}}$ and April 17$^{\text{th}}$, followed by opening of selected shops, re-starting of companies' activities and re-opening of all other shops. The percentage of the remained traffic (compared to the normal traffic) per week is shown in the centre of each fingerprint.

Comparing the characteristic fingerprints allows to disaggregate the overall traffic volume into separate intensities associated with different activities. This is realised in the lower part of Figure 8, which depicts the result of subtracting each weekly fingerprint in the top part of the figure from the fingerprint immediately on its left. This highlights better the time slots where traffic has increased (red) or decreased (green) compared to the previous fingerprint. For instance, since the introduction of the first relaxation measures on April 18$^{\text{th}}$, when people were allowed again to go to do-it-yourself and garden shops, traffic increased on average with 7% (3$^{\text{rd}}$ fingerprint on the bottom row). It is remarkable that this increase in traffic can also be observed on Sunday, when shops are closed. This seems to suggest that the traffic might not be exclusively for shopping, but that people perhaps stopped following travel restrictions strictly.

4.3 Insightful Blueprints of Brussels Traffic

The purpose of the following analysis is to get better insights of different traffic modes of the small ring of Brussels, which is actually a sequence of many tunnels crossing Brussels from one side to the other and is notoriously known for frequent traffic problems.

Traffic Profile Extraction Our analysis focuses on 16 locations of the small ring of Brussels. Our first aim is to identify characteristic traffic modes when considering all the 16 locations as one connected trajectory of tunnels. For this purpose, we perform the following steps:

- We consider only data from the weeks before the Covid-19 restrictions and exclude the carnival holidays. We order the data set in such a way that the locations are sequentially ordered as they spatially appear in reality.
- We associate with each timestamp a vector of 48 dimensions, based on the vehicle count, speed and road occupancy measurements that are available for each location.
- We average for each timestamp the measures over a rolling time window of 10 minutes, in order to achieve more resilient, but still fine-grained (per minute) results, Since this real-world data contains missing timestamps, this approach enables us to still use an estimation by taking the average of the non-missing values within the 10 minute time window.
- We scale each of the measures by the min-max normalisation [7]. This way all values will be between 0 and 1, resulting in equal weights during clustering.
- We cluster these vectors using k-means clustering using the Euclidean distance, resulting in clusters (of timestamps) representing characteristic traffic modes. The number of clusters was determined by majority voting of multiple validation measures, as explained in Section 3.1, resulting in 5 clusters.

In order to facilitate the semantic interpretation of the clusters (or traffic modes), we label them as follows: **Mode A:** Night traffic (avg. occupancy 7%); **Mode B:** Morning rush hour peak (avg. occupancy 26%); **Mode C:** Day traffic outside rush hours (avg. occupancy 22%); **Mode D:** Evening rush hour peak (avg. occupancy 28%); **Mode E:** Early morning and late evening traffic (avg. occupancy 15%). Remark that, in practice, occupancy is never close to 100% since this would require all vehicles to touch each other.

Temporal Traffic Blueprints In a second step, we use the traffic modes previously identified and visualise them in a label map. Figure 9 shows a label map as proposed in Section 3.4, which labels each timestamp using the 5 labels identified above. Note that the grey zones in the plot are caused by missing data stretching over more than the 10 minute time window.

This visualisation allows to observe the following interesting phenomena:

- Traffic during the first weeks of the lockdown was significantly reduced, resembling low intensity night traffic (mode A).
- During lockdown relaxations, traffic gradually evolved towards what is normally light traffic in the late evening and early morning (mode E)
- Day traffic outside rush hours (mode C) emerges in a similar way as mode E while lockdown measures are relaxed, initially starting at specific periods in the afternoon but slowly unrolling forward covering the whole afternoon.
- The afternoon rush hour (mode D) which can normally be observed between 15:00 and 18:00 completely disappeared during the lockdown and has since (almost) not yet reappeared.

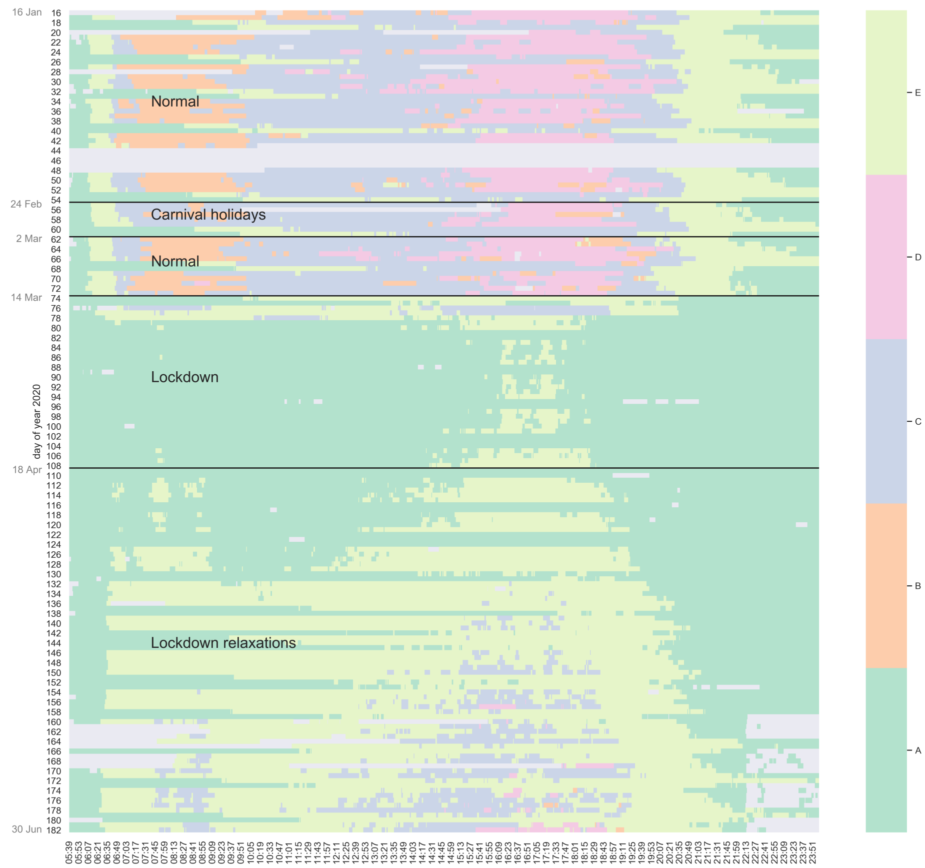


Fig. 9: Evolution of traffic clusters (modes) in Brussels.

5 Conclusion and Future Work

In this work, we proposed a range of visual analytics methods dedicated to spatio-temporal data. The power of the proposed visualisations lies in the transformation of data analytics results into well-thought visual representations revealing underlying insights. Questions as how to aggregate data, partition data, normalise data, represent it as a feature vector, etc. are the key to arrive at insightful visual representations facilitating the human eye's discovery process.

There are still many opportunities to expand this research, e.g. increase the intelligibility of visuals by making them interactive. Enabling to zoom in on time periods of interest or to change dynamically some parameters might be very valuable in supporting human-in-the-loop data exploration.

References

1. Micah Allen, Davide Poggiali, Kirstie Whitaker, Tom R Marshall, and Rogier Kievit. Raincloud plots. *PeerJ Preprints*, 6:e27137v1, 2018.
2. Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *symposium on visual analytics science and technology*. IEEE, 2008.
3. Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.
4. David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 2:224–227, 1979.
5. Julia Handl, Joshua Knowles, and Douglas B Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212, 2005.
6. David L Iverson. Inductive system health monitoring. *NASA*, 2004.
7. Zhenyu Liu et al. A method of svm with normalization in intrusion detection. *Procedia Environmental Sciences*, 11:256–262, 2011.
8. Yuyu Luo, Xuedi Qin, Nan Tang, and Guoliang Li. Deepeye: towards automatic data visualization. In *34th International Conf. on Data Engineering*. IEEE, 2018.
9. James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1.14, pages 281–297. Oakland, CA, USA, 1967.
10. Nils Rodrigues, Rudolf Netzel, Kazi Riaz Ullah, Michael Burch, Alexander Schultz, Bruno Burger, and Daniel Weiskopf. Visualization of time series data with spatial context. In *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction*, pages 37–44, 2017.
11. William M Spears. An overview of multidimensional visualization techniques. In *Evolutionary Computation Visualization Workshop*, 1999.
12. Guodao Sun, Ronghua Liang, Huamin Qu, and Yingcai Wu. Embedding spatio-temporal information into maps by route-zooming. *IEEE transactions on visualization and computer graphics*, 23(5):1506–1519, 2016.
13. Ying Tang, Fengfan Sheng, Hongxin Zhang, Chaojie Shi, Xujia Qin, and Jing Fan. Visual analysis of traffic data based on topic modeling (chinavis 2017). *Journal of Visualization*, 21(4):661–680, 2018.
14. J.J. Thomas and K.A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, Jan 2006.
15. Edward R Tufte, Nora Hillman Goeler, and Richard Benson. *Envisioning information*, volume 126. Graphics press Cheshire, CT, 1990.
16. Manasi Vartak, Silu Huang, Tarique Siddiqui, Samuel Madden, and Aditya Parameswaran. Towards visualization recommendation systems. *ACM SIGMOD Record*, 45(4):34–39, 2017.
17. Pak Chung Wong and Jim Thomas. Visual analytics. *IEEE Computer Graphics and Applications*, 1(5):20–21, 2004.
18. Jinfeng Zhao, Pip Forer, and Andrew S Harvey. Activities, ringmaps and geovisualization of large human movement fields. *Information visualization*, 7, 2008.